# ERATO 感謝祭 SeasonⅢ

9 & 10 August, 2016

Kawarabayashi Large Graph Project

## Day1, Tuesday 9 August 2016

### Makoto Yamada, Kyoto University

Beyond Ranking: Optimizing Whole-Page Presentation (WSDM'16)

Modern search engines aggregate results from different verticals: webpages, news, images, video, shopping, knowledge cards, local maps, etc. Unlike ten blue links", these search results are heterogeneous in nature and not even arranged in a list on the page. This revolution directly challenges the conventional ranked list" formulation in ad hoc search. Therefore, finding proper presentation for a gallery of heterogeneous results is critical for modern search engines. We propose a novel framework that learns the optimal page presentation to render heterogeneous results onto search result page (SERP).

### Yuichi Yoshida, National Institute of Informatics

Nonlinear Laplacian for Digraphs and its Applications to Network Analysis (WSDM'16)

In this work, we introduce a new Markov operator associated with a digraph, which we refer to as a nonlinear Lapla- cian. Unlike previous Laplacians for digraphs, the nonlinear Laplacian does not rely on the stationary distribution of the random walk process and is well defined on digraphs that are not strongly connected. We show that the nonlinear Laplacian has nontrivial eigenvalues and give a Cheeger-like inequality, which relates the conductance of a digraph and the smallest non-zero eigenvalue of its nonlinear Laplacian. Finally, we apply the nonlinear Laplacian to the analysis of real-world networks and obtain encouraging results.

### Takuya Konishi, National Institute of Informatics

Identifying Key Observers to Find Popular Information in Advance (IJCAI'16)

Identifying soon-to-be-popular items in web services offers important benefits. We attempt to identify users who can find prospective popular items. Such visionary users are called observers. We propose a feature selection based framework to identify efficient observers. This uses a binary classifier with sparse and non-negative constraints to predict item popularity. In experiments, we test our approach using real social bookmark datasets. The results demonstrate that our approach can find more efficient observers than baseline methods.

### Takahiro Ezaki, National Institute of Informatics

Reinforcement Learning Explains Conditional Cooperation and Its Moody Cousin (PLOS Computational Biology)

Behavioral experiments using repeated multiplayer social dilemma games on contact networks have revealed that players condition their decisions on the fraction of cooperative partners in the previous round of the repeated game. This behavior is referred to as the conditional cooperation. The origin of conditional cooperation and its moody variant still remains unclear. In this talk we provide a proximate explanation by numerical simulations. We show that players adopting a variant of the Bush-Mosteller reinforcement learning rule show the targeted behavior. We found that the reinforcement learners that showed (moody) conditional cooperation obeyed a behavioral

pattern corresponding to the GRIM strategy, a well-known strategy in the repeated prisoner's dilemma game.

### Yukino Baba, Kyoto University

Assessing Translation Ability through Vocabulary Ability Assessment (IJCAI'16)

We propose a practical method for assessing translation ability. Our key idea is to incorporate translators' vocabulary knowledge for translation ability assessment. Our method involves just asking translators to tell if they know given words. Using this vocabulary information, we build a probabilistic model to estimate the translators' vocabulary and translation abilities simultaneously. The results of our experiments show that the proposed method accurately estimates translation ability and selects translators who have sufficient skills in translating a given sentence.

### Ryusuke Takahama, Kyoto University

Progressive Comparison for Ranking Estimation (IJCAI'16)

Object ranking is a problem that involves ordering given objects by aggregating pairwise comparison data collected from one or more evaluators; however, the cost of object evaluations is high in some applications. In this paper, we propose an efficient data collection method called progressive comparison, whose objective is to collect many pairwise comparison data while reducing the number of evaluations. We also propose active learning methods to determine which object should be evaluated next in the progressive comparison; we propose two measures of expected model changes, one considering the changes in the evaluation score distributions and the other considering the changes in the winning probabilities. The results of experiments using a synthetic dataset and two real datasets demonstrate that the progressive comparison method achieves high estimation accuracy with a smaller number of evaluations than the standard pairwise comparison method, and that the active learning methods further reduce the number of evaluations as compared with a random sampling method.

### Daisuke Hatano, National Institute of Informatics

Adaptive Budget Allocation for Maximizing Influence of Advertisements (IJCAI'16)

The budget allocation problem is an optimization problem arising from advertising planning. In the problem, an advertiser has limited budgets to allocate across media, and seeks to optimize the allocation such that the largest fraction of customers can be influenced. It is known that this problem admits a $(1 - 1/e)$-approximation algorithm. However, no previous studies on this problem considered adjusting the allocation adaptively based upon the effect of the past campaigns, which is a usual strategy in the real setting. Our main contribution in this paper is to analyze adaptive strategies for the budget allocation problem. We define a greedy strategy, referred to as the insensitive policy, and then give a provable performance guarantee. This result is obtained by extending the adaptive submodularity, which is a concept studied in the context of active learning and stochastic optimization, to the functions over an integer lattice.

### Yasushi Sakurai, Kumamoto University

Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences (KDD'16)

Mining and Forecasting of Big Time-series Data (SIGMOD'15)

Non-Linear Mining of Competing Local Activities (WWW'16)

Mining Big Time-series Data on the Web (WWW'16)

Given a large collection of time series, such as web-click logs, electric medical records and motion capture

sensors, how can we efficiently and effectively find typical patterns? How can we statistically summarize all the sequences, and achieve a meaningful segmentation? What are the major tools for forecasting and outlier detection? Time-series data analysis is becoming of increasingly high importance, thanks to the decreasing cost of hardware and the increasing on-line processing capability. The objective of this tutorial is to provide a concise and intuitive overview of the most important tools that can help us find patterns in large-scale time-series sequences. We review the state of the art in four related
fields: (1) similarity search and pattern discovery, (2) linear modeling and summarization, (3) non-linear modeling and forecasting, and (4) the extension of time-series mining and tensor analysis. The emphasis of the tutorial is to provide the intuition behind these powerful tools, which is usually lost in the technical literature, as well as to introduce case studies that illustrate their practical use.

### Ichiro Takeuchi, Nagoya Institute of Technology

Safe Pattern Pruning: An Efficient Approach for Predictive Pattern Mining (KDD'16)

In this paper we study predictive pattern mining problems where the goal is to construct a predictive model based on a subset of predictive patterns in the database. Our main contribution is to introduce a novel method called safe pattern pruning (SPP) for a class of predictive pattern mining problems. The SPP method allows us to efficiently find a superset of all the predictive patterns in the database that are needed for the optimal predictive model. The advantage of the SPP method over existing boosting-type method is that the former can find the superset by a single search over the database, while the latter requires multiple searches. The SPP method is inspired by recent development of safe feature screening. In order to extend the idea of safe feature screening into predictive pattern mining, we derive a novel pruning rule called safe pattern pruning (SPP) rule that can be used for searching over the tree defined among patterns in the database. The SPP rule has a property that, if a node corresponding to a pattern in the database is pruned out by the SPP rule, then it is guaranteed that all the patterns corresponding to its descendant nodes are never needed for the optimal predictive model. We apply the SPP method to graph mining and item-set mining problems, and demonstrate its computational advantage.

### Yasuo Tabei, Tokyo Institute of Technology

Scalable Partial Least Squares Regression on Grammar-Compressed Data Matrices (KDD'16)

We present a novel compressed representation of data matrices for learning statistical models with high interpretability.
Experiments using various massive high-dimensional data show that our method enables us to space-efficiently learn the linear model in PLS on the compressed representation of data matrices and our PLS performs superiorly in terms of prediction accuracy, computational efficiency, and interpretability.

### Takuya Akiba, Preferred Networks

Compact and Scalable Graph Neighborhood Sketching (KDD'16)

The all-distances sketch (ADS) has recently emerged as a promising paradigm of graph neighborhood sketching. An ADS is a probabilistic data structure that is defined for each vertex of a graph. ADSs facilitate accurate estimation of many useful indicators for network analysis with the guarantee of accuracy, and the ADSs for all the vertices in a graph can be computed in near-linear time. Because of these useful properties, ADS has attracted considerable attention. However, a critical drawback of ADS is its space requirement, which tends to be much larger than that of the graph itself. In the present study, we address this issue by designing a new graph sketching

scheme, namely, sketch retrieval shortcuts (SRS). Although SRSs are more space-efficient than ADSs by an order of magnitude, an ADS of any vertex can be quickly retrieved from the SRSs. The retrieved ADSs can be used to estimate the aforementioned indicators in exactly the same manner as with plain ADSs, inheriting the same accuracy guarantee. Our experiments on real-world networks demonstrate the usefulness of SRSs as a practical back-end of large-scale graph data mining.

## Day2, Wednesday 10 August 2016

### Tasuku Soma, The University of Tokyo

A Generalization of Submodular Cover via the Diminishing Return Property on the Integer Lattice (NIPS'15)

We consider a generalization of the submodular cover problem based on the concept of diminishing return property on the integer lattice. We are motivated by real scenarios in machine learning that cannot be captured by (traditional) submodular set functions. We show that the generalized submodular cover problem can be applied to various problems and devise a bicriteria approximation algorithm. Our algorithm is guaranteed to output a log-factor approximate solution that satisfies the constraints with the desired accuracy. The running time of our algorithm is roughly O(n¥log (nr) ¥log{r}), where n is the size of the ground set and r is the maximum value of a coordinate. The dependency on r is exponentially better than the naive reduction algorithms. Several experiments on real and artificial datasets demonstrate that the solution quality of our algorithm is comparable to naive algorithms, while the running time is several orders of magnitude faster. This is joint work with Yuichi Yoshida.

### Masaaki Imaizumi, The University of Tokyo

Doubly Decomposing Nonparametric Tensor Regression (ICML'16)

Nonparametric extension of tensor regression is proposed. Nonlinearity in a high-dimensional tensor space is broken into simple local functions by incorporating low-rank tensor decomposition. Compared to naive nonparametric approaches, our formulation considerably improves the convergence rate of estimation while maintaining consistency with the same function class under specific conditions. To estimate local functions, we develop a Bayesian estimator with the Gaussian process prior. Experimental results show its theoretical properties and high performance in terms of predicting a summary statistic of a real complex network.

### Song Liu, The Institute of Statistical Mathematics

Structure Learning of Partitioned Markov Networks (ICML'16)

We learn the structure of a Markov Network between two groups of random variables from joint observations. Since modelling and learning the full MN structure may be hard, learning the links between two groups directly may be a preferable option. We introduce a novel concept called the partitioned ratio whose factorization directly associates with the Markovian properties of random variables across two groups. A simple one-shot convex optimization procedure is proposed for learning the sparse factorizations of the partitioned ratio and it is theoretically guaranteed to recover the correct inter-group structure under mild conditions. The performance of the proposed method is experimentally compared with the state of the art MN structure learning methods using ROC curves. Real applications on analyzing bipartisanship in US congress and pairwise DNA/time-series alignments are also reported.

### Takanori Maehara, Shizuoka University

Joint word representation learning using a corpus and a semantic lexicon (AAAI'16)

Methods for learning word representations using large text corpora have received much attention lately due to their impressive performance in numerous natural language processing (NLP) tasks such as, semantic similarity measurement, and word analogy detection.

Despite their success, these datadriven word representation learning methods do not consider the rich semantic

relational structure between words in a cooccurring context.
On the other hand, already much manual effort has gone into the construction of semantic lexicons such as the WordNet that represent the meanings of words by defining the various relationships that exist among the words in a language.
We consider the question, can we improve the word representations learnt using a corpora by integrating the knowledge from semantic lexicons?
For this purpose, we propose a joint word representation learning method that simultaneously predicts the co-occurrences of two words in a sentence subject to the relational constrains given by the semantic lexicon.
We use relations that exist between words in the lexicon to regularize the word representations learnt from the corpus.
Our proposed method statistically significantly outperforms previously proposed methods for incorporating semantic lexicons into word representations on several benchmark datasets for semantic similarity and word analogy.

**Kohei Hayashi,** National Institute of Advanced Industrial Science and Technology

Expected Tensor Decomposition with Stochastic Gradient Descent (AAAI'16)

In this study, we investigate expected CP decomposition — a special case of CP decomposition in which a tensor to be decomposed is given as the sum or average of tensor samples. To determine this decomposition, we develope stochastic-gradient-descent-type algorithms with four appealing features: efficient memory use, ability to work in an online setting, robustness of parameter tuning, and simplicity. Our theoretical analysis show that the solutions do not diverge to infinity for any initial value or step size. Experimental results confirm that our algorithms significantly outperform all existing methods in terms of accuracy. We also show that they can successfully decompose a large tensor, containing billion-scale nonzero elements.

**Kenji Fukumizu,** The Institute of Statistical Mathematics

Persistence weighted Gaussian kernel for topological data analysis (ICML'16)

Topological data analysis (TDA) is an emerging mathematical method for extracting topological and geometrical information of data. In the method, a persistence diagram (PD), 2-D plot for illustrating the topology, is widely used as a descriptor of data.
In this work, we introduce a kernel method for PD to apply statistical methods systematically to TDA, aiming at flexibly distinguishing significant topological structures from noise. We also discuss a computational challenge of TDA, and propose an approximate computation method. As a theoretical background, a stability theorem is proved. The proposed kernel is applied to practical data analysis in materials science and biology, showing favorable results in comparison with other existing methods.

**Mathieu Blondel,** NTT Communication Science Laboratories

Polynomial Networks and Factorization Machines: New Insights and Efficient Training Algorithms (ICML'16)

Polynomial networks and factorization machines are two recently-proposed models that can efficiently use feature interactions in classification and regression tasks. In this paper, we revisit both models from a unified perspective. Based on this new view, we study the properties of both models and propose new efficient training algorithms. Key to our approach is to cast parameter learning as a low-rank symmetric tensor estimation problem, which we solve by multi-convex optimization. We demonstrate our approach on regression and recommender

system tasks.

### Atsushi Miyauchi, Tokyo Institute of Technology

What Is a Network Community? A Novel Quality Function and Detection Algorithms (CIKM'15)

We introduce a novel quality function for a network community, which we refer to as the communitude. Specifically, it measures the Z-score of a subset of vertices S with respect to the fraction of the number of edges within the subgraph induced by S. To evaluate the detection ability of our quality function, we address the communitude maximization problem and its variants for realistic scenarios. For the problems, we propose a linear-time heuristic algorithm together with some modified versions. Computational experiments using both synthetic graphs and real-world networks demonstrate the validity of the proposed quality function and algorithms.

### Naoto Ohsaka, The University of Tokyo

Dynamic Influence Analysis in Evolving Networks (VLDB'16)

We propose the first real-time fully-dynamic index data structure designed for influence analysis on evolving networks. With this aim, we carefully redesign the data structure of the state-of-the-art sketching method introduced by Borgs et al., and construct corresponding update algorithms. Using this index, we present algorithms for two kinds of queries, influence estimation and influence maximization, which are strongly motivated by practical applications, such as viral marketing. We provide a thorough theoretical analysis, which guarantees the non-degeneracy of the solution accuracy after an arbitrary number of updates. Furthermore, we introduce a reachability-tree-based technique and a skipping method, which greatly reduce the time consumption required for edge/vertex deletions and vertex additions, respectively, and counter-based random number generators, which improve the space efficiency.

Experimental evaluations using real dynamic networks with tens of millions of edges demonstrate the efficiency, scalability, and accuracy of our proposed indexing scheme. Specifically, it can reflect a graph modification within a time of several orders of magnitude smaller than that required to reconstruct an index from scratch, estimate the influence spread of a vertex set accurately within a millisecond, and select highly influential vertices at least ten times faster than state-of-the-art static algorithms.

### Naonori Kakimura, The University of Tokyo

Exact and Approximation Algorithms for Weighted Matroid Intersection (SODA'16)

The weighted matroid intersection is a common generalization of various combinatorial optimization problems such as bipartite matchings, packing spanning trees, and arborescences in a directed graph. In this talk, we propose new exact and approximation algorithms for the weighted matroid intersection problem. Our exact algorithm is faster than previous algorithms when the largest weight is relatively small. Our approximation algorithm delivers a (1-epsilon)-approximate solution with a running time significantly faster than most known exact algorithms.

### Shuichi Hirahara, The University of Tokyo

Limits of Minimum Circuit Size Problem as Oracle (CCC'16)

The Minimum Circuit Size Problem (MCSP) is known to be harder than NP-intermediate problems such as integer factorization. However, it is one of important open problems in computational complexity whether MCSP is NP-hard or not. In this talk, we prove that the current reduction techniques cannot establish NP-hardness of MCSP. Specifically, we introduce the notion of oracle-independent reduction, which captures current reduction techniques,

and then we prove that oracle-independent reductions have certain inherent limits.

Akitoshi Kawamura, The University of Tokyo

A lower bound on opaque sets (SoCG'16)

It is proved that the total length of any set of countably many rectifiable curves, whose union meets all straight lines that intersect the unit square U, is at least 2.00002. The best known lower bound has been 2. A similar bound is proved for all convex sets U other than a triangle.